引用格式：李爱华, 续维佳, 石勇. 基于“物理—事理—人理”的多源异构大数据融合探究. 中国科学院院刊, 2023, 38(8): 1225-1233, doi: 10.16418/j.issn.1000-3045.20220921003.
Li A H, Xu W J, Shi Y. Research on multi-source heterogeneous big data fusion based on WSR. Bulletin of Chinese Academy of Sciences, 2023, 38(8): 1225-1233, doi: 10.16418/j.issn.1000-3045.20220921003. (in Chinese)

基于“物理—事理—人理”的多源异构大数据融合探究

李爱华[1] 续维佳[1] 石 勇[2,3,4*]

1 中国财经大学 管理科学与工程学院 北京 100081
2 中国科学院大学 经济与管理学院 北京 100190
3 中国科学院虚拟经济与数据科学研究中心 北京 100190
4 中国科学院大数据挖掘与知识管理重点实验室 北京 100190

摘要 多源异构大数据时代下，大数据呈现出交叉性、多元性、变化性等新特征，更广泛领域的应用对数据融合产生新需求，在此背景下数据融合的内涵得到丰富和扩展。广义的数据融合包含对数据资源的融合、模型方法的融合及决策者知识和经验的融合。文章分析了多源异构数据融合在数据层、信息层和决策层3个不同融合层次的特点，探讨了数据融合在存储、使用、分析技术、数据管理及价值确定方面可能面临的挑战，并提出了相应的对策建议，为企业、政府等各类主体高效管理数据资源，进行更深入的数据融合分析提供参考。

关键词 多源异构，物理—事理—人理（WSR），数据融合，大数据

DOI 10.16418/j.issn.1000-3045.20220921003
CSTR 32128.14.CASbulletin.20220921003

在大数据时代，海量数据作为各领域的重要资源，其更多的价值有待被挖掘，更大的潜能有待释放。此外，移动通信设备的深度普及、互联网的全面覆盖，以及5G、人工智能（AI）等的高速发展，不断推动大数据自身的更新和前进。从企业产生的内部数据在生产经营等环节中蕴含着重要价值，到个人产生的数据在健康管理及智慧医疗等应用场景发挥着重要作用；从海量数据采集困难重重到社会经济活动随时

*通信作者
资助项目：国家自然科学基金重点项目（71932008、72231010）
修改稿收到日期：2023年6月25日

随地产生数据司空见惯，数字已赋能全社会、各领域、各行业产业的发展。

在科学技术进步和社会发展变革的背景下，当前大数据除了拥有巨大的体量，多源异构的特点也充分凸显，对海量数据的分析也产生了许多新需求和新挑战。因此，如何科学、有效、全面地融合多种来源、各种类型的“硬数据”和“软数据”，化解多源异构大数据融合在理论和实践中的难题，使数据的堆砌变为具有价值的“黄金”，成为有待解决的重要问题。

1 多源异构大数据新特征与数据融合新需求

互联网连接了人们的生活、企业的生产和政府的管理，也使得社会各参与者在无数活动中产生数据。这些数据来源广泛，结构复杂多样，同时，数据可获得性的增强也使更多领域日益重视数据资源价值的挖掘。因此，海量数据的新特征及更多领域的新需求，促使多源异构大数据融合成为大数据研究领域的重要内容。

多源异构大数据呈现的新特征总体可以概括为：交叉性、多元性、变化性和共识性。不同活动、不同业务的内容常常出现重叠，大量跨行业、跨媒体、跨数据库的数据具有很强的交叉性。数据形态结构也呈现多样化，既有数字、表格等结构化数据，也有文本、图片、声音、视频等非结构化及半结构化数据。此外，多元性不仅是数据类型、结构不同，而且是数据中蕴含的内容和知识的多“维度”与多“粒度”，体现了数据和知识间复杂的立体关系[1]。数据的变化性指数据随着时间的推移发生变化。数据的共识性是指人们对很多数据之间的关系及数据和知识之间的关系已经达成共识，这些关系和知识具有普遍适用性，有助于建立起数据、信息和知识间的关联关系，挖掘更多知识。

挖掘多源异构数据中的信息和知识并将其转化为价值，离不开数据融合。过去的数据融合也称为信息融合，其内涵为组合和处理多来源的数据和信息，使其优势互补、排除噪声、化解矛盾，提高信息的完整性和可信度，得到比单一信息更准确、更可靠的估计或决策[2]。信息融合的模型主要包括结构模型和功能模型。结构模型描述了信息融合系统的工作方式，在融合功能的部署上分为集中式结构、分布式结构和混合式结构[3]。功能模型主要对信息融合系统及子系统的功能作用及各部分的关系进行建模，包括JDL模型（Joint Directors of Laboratories）[4]、Omnibus模型[5]、OODA模型（Observation, Orientation, Decision, Action）及其改进模型[6,7]。JDL改进模型[8]对多元信息融合任务构建了6级功能：次目标估计、目标评估、态势评估、影响评估、过程优化、认知优化。OODA模型由观测、定向、决策和行动4部分组成，其扩展模型可以处理相互影响的信息之间的融合[6]。信息融合模式可以抽象为3个层级，分别是数据级融合、特征级融合和决策级融合[3]。数据级融合主要指对相同介质传感器收集的数据直接进行融合；特征级融合是指先从原始数据中提取特征，然后对它们进行融合；决策级融合则是对特征或已经得到的初步结果进行更高层次的融合，得出更综合和系统的决策。不同融合层次都有大量方法和技术，包括主成分分析、Kalman滤波、贝叶斯估计、机器学习、D-S证据理论、智能计算等。从数据级融合到特征级融合再到决策级融合，融合的层次越来越高，计算量和精度降低，信息损失增加，但稳健性和灵活性得到提升。多元信息融合的应用主要在军事[9]、通信[10]、控制[11]、医疗[12]、交通[13]等领域，融合的数据以传感器产生的“硬数据”为主。

多源异构大数据对数据融合的新需求使数据融合在理论研究、方法技术和实际应用中都迎来了挑战。当前由于多源异构大数据在全领域、各行业的渗透和价值的发挥，数据挖掘与融合的应用延伸到企业管理、政府治理、银行风险防范等诸多社会经济问题中，这些新兴的应用场景有别于信息融合传统的应用

领域，因此，企业、政府等众多参与主体的应用场景对数据融合提出了新需求。在社会经济和管理领域中，数据来源更具复杂性和开放性，研究问题往往参与主体较多，系统性较强，存在大量直接关联或隐性关联的数据，与传统传感器收集的数据相比，预设数据源的重要性可能降低，开发和确定新数据变得重要。同时，社会活动离不开“人”的参与，因此充分考虑与“人”有关的“软因素”和“软数据”，将其与其他“硬数据”融合分析，成为新的需求。在数据融合模式上，将多个层级的融合相结合而非单独局限于某个层次的融合，使数据融合贯穿于数据挖掘全过程成为未来发展方向。此外，社会、经济和管理领域的应用场景需要强化前沿技术与领域专家知识的结合，增强方法工具在应用中的可理解性及和实践的联系具有重要的意义。

2 基于WSR的多源异构大数据融合研究思路

如前所述，鉴于当前跨媒体、跨行业等多源异构数据的交叉性、多元性、变化性和共识性，要实现对形态结构各异的数据进行统一分析和挖掘，离不开数据融合。李爱华等[14]对比分析了信息融合的3个层级和商务智能中“数据、信息、知识”3个层次的关联和区别，以此为基础，借鉴“物理—事理—人理”（WSR）系统科学方法论的思想，提出了广义数据融合的内容，将其贯穿于商务智能的分析过程。

WSR系统科学方法论[15]综合考虑了系统实践活动中“物”“事”“人”3个维度，强调了客观世界、系统组织和人的动态统一和紧密联系，它们三者构成整体，缺一不可[16,17]。“物理”是对现实世界的组成、属性和客观规律的研究，“事理”是解决问题的方法，“人理”是对人的动态活动、思想行为及和环境的相互影响等的研究。在社会、经济和管理等领域，人是各种活动的重要参与者，实际问题的解决和决策制定有赖于客观状况、解决方式，以及与人相关的因素等多个方面，这与WSR方法论中“物”“事”和“人”形成对应。基于WSR，本文提出的广义数据融合的内涵是：综合运用多种方法，对多源异构的原始数据进行挖掘，将得到的内涵、模式、决策及其他“软因素”等进行综合、全面地处理和分析[14]，最终实现高效的融合效果，为决策提供辅助和支持。广义多源异构大数据融合包含数据资源的融合、模型方法的融合和决策者知识和经验的融合。在商务智能中，“数据”是通过各种渠道获取的、未经加工的原始资源；“信息”是对数据初步分析后挖掘得到的潜在特征、关联和规律等；“知识”是通过进一步推理得到的更有价值的结论等。数据为解决问题提供了原材料，信息和知识为决策制定提供了依据和支持。由于基于WSR方法论的广义数据融合贯穿于商务智能分析“数据—信息—知识”的全过程，因此可以划分为数据层融合、信息层融合和知识层融合3个层次[18]（图1）。

WSR在多源异构数据层融合主要体现在数据源的确定和数据的收集。在社会治理、企业管理、经济发展、风险管理等问题场景中，由移动设备、网络等收集的人的行为活动数据发挥着越来越重要的作用；研究问题的跨行业、跨领域、跨学科的交叉性，以及多源异构数据的跨媒体性，大大提升了数据源的数量，同时也增加了数据源选择和确定的难度。数据的选择具有一定的主观性，针对同样的问题，选择不同的数据可能有不同角度的发现。数据的选取需要综合考虑实际问题和领域专家经验[19]。因此，基于WSR可以从“物”“事”“人”3个维度出发对数据进行选择，即客观数据、行为活动产生的数据和评价、观点、情绪、判断、预期等与“人”密切相关的数据。此外，多源异构数据转化和综合性指标的构建也是数据层融合的内容。结构化、半结构化、非结构化、不同粗细粒度、动态、静态的数据很难直接进行建模分析。因此，需要对多源数据进行转化，通过聚合、关联、特征提取、文本挖掘、计算新变量等多种方法技术使它

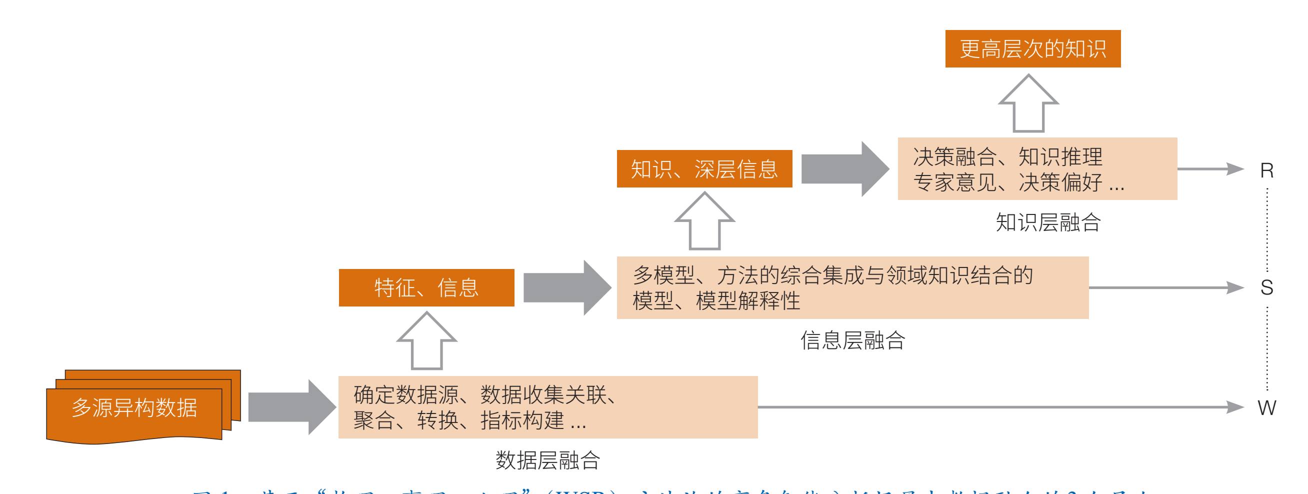


图 1　基于“物理—事理—人理”（WSR）方法论的商务智能分析场景中数据融合的 3 个层次

Figure1　Three levels of data fusion in business intelligence analysis scenarios based on WSR

们能够进行统一分析，为信息层融合构建综合性模型和挖掘深层次信息提供基础。WSR 中的“人理”表现为数据转换和指标建立过程中的可理解性与实际意义。

WSR 在数据融合信息层的应用不仅体现在将多种模型方法进行综合集成，对数据层融合结果进行分析，还体现在人们对模型方法的选择，以及数据科学技术与社会、经济等领域知识、原理和方法的结合。社会治理、经济发展、企业管理领域有自身特点和基本的理论、原理，数据融合的技术工具不能脱离原有学科领域的理论基石。因此，在方法的选择上需要不断探索传统方法与新兴技术互相结合的路径，寻求结果准确性和可解释性之间的平衡。信息层融合“事理”中模型建立的常用方法包括分类、聚类、关联规则挖掘等机器学习、深度学习和人工智能方法，以及多种方法的集成模型等。此外，数据融合信息层中方法的选择与“人理”密切相关，在解决实际的社会管理问题时，需要在全面考虑有关参与各方的基础上，采用合适的方法和模型，而非一味地追求方法的复杂和结果的精确。

WSR 在知识层融合中的应用主要体现在最终的决策需要有机结合“人理”，将与人有关的因素与客观数据分析结果相融合仍然为研究的热点。知识层融合的“物理”是从信息层融合得到的结果，通过结合专家意见、决策者偏好等更多因素，通过采用更高层次的推理、挖掘方法，将低阶段融合结果再次进行融合分析，得到深度关系和可领悟的知识，实现需求的满足或决策的支持。决策的制定和知识的认知与人密切相关，专家意见、决策者偏好、社会环境等都可能影响最终的结果，因此“人理”在数据融合知识层融合中的重要性不可轻视。

3 多源异构大数据融合在数据应用环节中遇到的问题和挑战

在新应用场景中，多源异构大数据融合产生的新特点促使其内涵得到扩展和延伸，同时也给多源异构大数据融合在存储、集成、分析、管理等方面带来新的挑战和难题。

*（1）高质量的数据存储问题亟待解决。*数据存储是数据分析的基础和前端工作，数据存储越有效，后续的数据提取、预处理和分析等将越便捷和高效，因此数据存储在整个融合分析过程中十分重要。然而，数据存储工作本身繁琐冗杂，当前爆炸式增长数据的多源异构性更给数据存储增加了难度。面对繁杂的数

据来源，数据存储要解决的问题是：① 存储哪些数据。对历史数据的清理和去除可以节约存储空间，但也可能流失有价值的数据资源，因此需要权衡数据重要性和空间占用的关系。② 如何存储数据。针对多源异构大数据，如何条理清晰存储形态各异的数据成为不可避免的问题。存储的数据质量也至关重要。数据质量显著影响分析的结果，如果不重视存储阶段，导致数据保存不规范甚至出现错误，会严重影响后续分析的效率和准确性。

（2）数据孤岛与使用壁垒导致数据集成融合困难。虽然每时每刻都有大量数据产生，但真正有效使用这些数据仍面临困难，大部分数据以“数据孤岛”的形式存在，彼此之间互不连通，数据的利用存在明显的壁垒。数据层的融合与集成难度有2个方面：① 来源于主观因素，即数据的使用权限，很多内部数据不对外开放，难以获取；② 来源于客观因素，即数据由不同业务活动产生，即使在同一个企业或机构内部，数据也由于碎片化问题、不同的结构形式和存储标准等，导致跨部门数据的使用困难。这些因素增加了数据资源融合的难度，不利于充分挖掘数据价值。

（3）多源、异构大数据增加了不同场景下数据分析与挖掘的难度。多源异构大数据给数据融合技术带来许多新挑战。数据融合不仅要实现多源异构数据的转化和统一集成分析，还需要关注数据背后的隐含知识，强化对数据意义的理解，将共识性的知识与数字的分析有机结合。在融合模型和方法中，解决跨领域、跨媒体、跨语言、多学科的融合仍然为研究的前沿和难点。数据融合的对象以数字、表格、文字、图片、视频、音频，或者知识、规律、模型等多形式共存，内容涉及不同领域，甚至可能包含不同语言，需要充分考虑不同数据资源的特点和不同领域的差异及共性，跨语言的融合则有赖于跨语言数据关联和大规模知识库。此外，当前海量、多源、异构数据也对数据处理和分析速度提出新的要求，如很多价值蕴藏在高频数据或者数据流中，需要高效快速地实时处理数据技术；同时，数据融合方法和技术由于数据量的增加也需要不断优化。

（4）数据的维护、安全和隐私泄漏是当前数据管理需要关注的重点。多源异构大数据需要高性能的网络架构和强大的数据中心支撑，因此数据仓库和数据中心的运营维护也将成为挑战。数据量庞大、动态演化使数据库及知识库的增量更新、错误恢复等操作难度大大增加，如何保证数据稳定、支持高并发的同时减少服务器的低负载情况也成为数据中心维护的重点内容。数据融合分析中，需要提高对数据安全的重视程度。硬件设备设施的故障、网络黑客的攻击等都可能导致数据资源的丢失。因此，对数据的多副本与容灾机制的建立需要不断加强。此外，信息安全问题也得到广泛关注，人们对隐私的保护日益重视。数据融合增强了数据间关联性，也使个人隐私、企业、国家安全信息等面临更大的泄漏风险和威胁，因此在分析过程中如何保护敏感信息，如何在灵活利用数据时保障数据安全也是数据融合未来研究中重要的课题。

（5）数据开放与共享、数据交换以及数据资产定价需要进一步关注。数据潜在价值的发挥与数据的开放程度密切相关，往往开放程度越高的数据，越能被挖掘出更多价值，应用于更多场景和领域。然而，数据的开放面临很多复杂问题。① 由于商业利益、行业垄断、信息安全等问题，数据的开放受到极大的限制。② 对数据权责清晰界定存在一定困难，例如个人用户常常是数据的生产者，也是数据的受益人。实际中数据所有者和权利往往不断发生变化，所有者及其权利的界定尚未达成明确的共识。③ 缺乏数据共享相关的完善的政策法规也制约了数据的开放。数据的价值越来越受到重视，数据的交换、交易及相关市场随之产生，如何界定数据交易价值，维护交易行为的安全和良性发展，保障个人、组织和国家的合法权益，成为多源异构大数据时代的新挑战。

4 多源异构大数据融合发展的思考

对于多源异构大数据，从复杂系统的研究视角来说，数据是客观存在的“物理”、融合不同数据的方法则是“事理”、对多源异构大数据的管理则和“人理”联系密切。因此，针对多源异构数据在存储、使用、分析、维护等方面存在的挑战，借鉴WSR3个维度协调统一的思想，从数据、方法和管理3个方面针对数据融合在应用中的挑战提出了3点思考（图2）。

（1）从数据角度上看，继续优化采集和存储。对于数据巨大、结构类型复杂的数据而言，存储和数据库建设是复杂的工程。首先要明确业务需求，发挥数据工程师和领域专家、业务人员的共同作用，开发大数据应用场景，这有赖于对客观数据本身特点（也即“物理”）的深入分析，对“人理”的充分理解，以及数据和人们需求之间的联系与协调。数据的存储不能仅局限于当前的需求，由于技术的进步和业务的更新，新的需求和潜在需求会不断产生，可以基于WSR的“物”“事”“人”3个维度确定数据存储资源。此外，对跨媒体多源异构大数据的收集和存储需要更先进的“事理”。在数据融合的新需求下，要进一步强化数据库的建设和维护，在数据存储时考虑数据的多源异构性，实现对结构化数据、半结构化和非结构化数据的兼容，建立数据融合溯源机制，从而提高数据库增量更新和局部修改的灵活性与简便性。高效、高质量的数据存储是大数据融合分析的基石，数据的存储需要最大限度地为数据的使用提供便利，清晰的格式、统一的标准等有利于高效的数据调用、处理、分析、更新和维护等，能够极大地节约资源和成本。

（2）从方法上看，需要多维度提升数据融合效果。提升海量多源异构数据的融合效果，有赖于硬件设备和技术的共同进步。复杂结构、数量庞大数据和信息的融合对硬件设备提出了更高的要求，提高硬件设备的性能，完善相关基础设施的建设，能够为未来进一步大数据融合的发展奠定坚实的基础。在融合方法方面，不断进行技术创新，针对数据层融合、信息层融合和决策层融合各层的特点、区别和需求，对原有算法、模型等进行改进、集成和融合。充分借鉴多学科的思维，从不同角度获取处理多源数据、融合多元知识的启示。此外，加强交叉学科人才的培养，在数据融合的理论研究和实际运用中，充分发挥数据科学家、领域专家、领域知识库的共同作用，形成“1+1>2”的互补优势。

（3）从管理角度看，需要建立共享机制，保障数

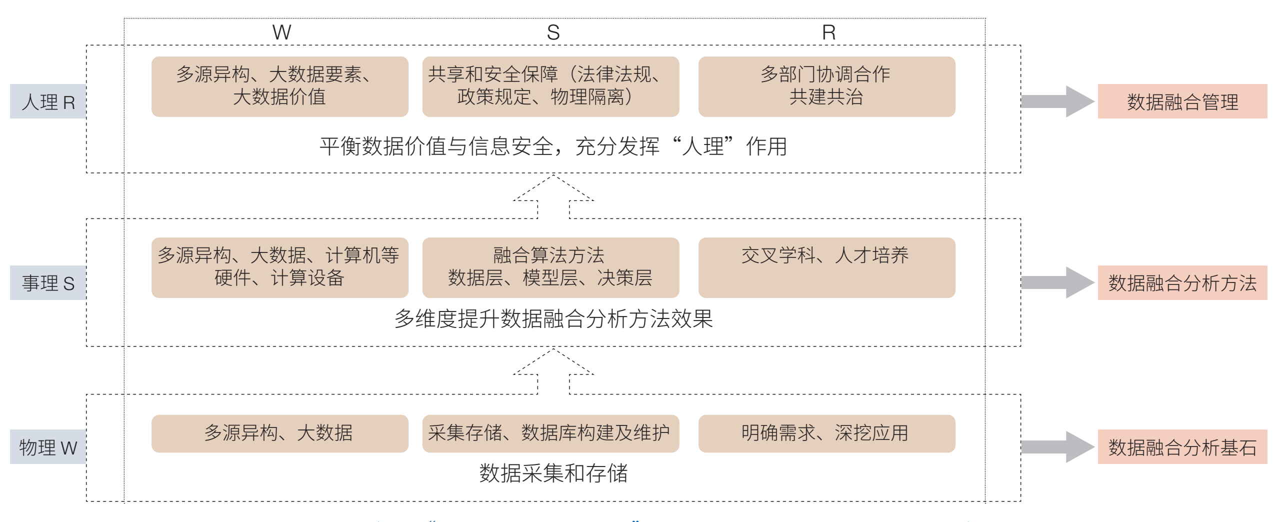


图2 基于“物理—事理—人理”（WSR）的数据融合发展的思考

Figure 2 Thinking on development of data fusion based on WSR

据开放和数据安全。当前数据作为一种新兴的要素，能够产生越来越多的价值，无论企业还是政府，都日益提高对数据的重视程度，不断增强大数据管理，提出与时俱进的数字化发展战略。因此，如何充分、高效、安全的实现数据价值成为重要的问题。数据价值的发挥和潜能的释放离不开数据的开放和共享，而数据的开放不可避免影响数据的安全。因此，需要全面、综合考虑各方利益，建立健全数据共享机制，不断完善相关法规政策，为数据共享、数据安全等提供强有力的法规保障，实现数据共享的同时遏制数据滥用，建立可持续发展的良性数据共享生态系统。保护数据与信息的安全还可以采取物理隔离与权限控制相结合的方法，通过隔离防止非法访问；研究降低隐私泄漏风险的策略和评估模型，及时进行风险预警和保护策略更新；强化大数据网络安全的构建等。在多源异构数据融合管理中，“人理”至关重要，连接数据孤岛、打破数据壁垒，离不开各个部门之间高效的沟通和协同合作。大数据共享生态系统需要全社会各主体参与共建共治，才能实现数据共享、利益保护、安全保障的良性环境，为未来数据融合的发展和数据价值的增长提供基石。

5 结语

在新兴应用场景下，多源异构大数据融合在数据层、信息层和知识层中有了新的特点和内涵，借鉴WSR系统科学方法论，综合物、事、人3个维度对数据融合的各个层次进行分析和研究，有利于更好地解决多源异构数据融合问题，为决策提供更具综合性的支持。数据融合对人类驾驭数据的能力提出新挑战，使数据在存储、使用、管理等多个方面产生了新的难题，但也为人们获得更为深刻、系统和综合的洞察能力以及更充分的数据价值挖掘和利用，提供了巨大的空间与潜力。

参考文献

1 孟小峰, 杜治娟. 大数据融合研究:问题与挑战. 计算机研究与发展, 2016, 53(2): 231-246.
Meng X F, Du Z J. Research on the big data fusion: Issues and challenges. Journal of Computer Research and Development, 2016, 53(2): 231-246. (in Chinese)

2 陈科文, 张祖平, 龙军. 多源信息融合关键问题、研究进展与新动向. 计算机科学, 2013, 40(8): 6-13.
Chen K W, Zhang Z P, Long J. Multisource information fusion: Key issues, research progress and new trends. Computer Science, 2013, 40(8): 6-13. (in Chinese)

3 彭冬亮, 文成林, 薛安克. 多传感器多源信息融合理论及应用. 北京: 科学出版社, 2010.
Peng D L, Weng C L, Xue A K. Theory and application of multi-sensor and multi-source information fusion. Beijing: Science Press, 2010. (in Chinese)

4 White J F E. A model for data fusion// Proceedings of the First National Symposium on Sensor Fusion. 1988, 2: 149-158.

5 Bedworth M, O’Brien J. The Omnibus model: A new model of data fusion?. IEEE Aerospace and Electronic Systems Magazine, 2000, 15(4): 30-36.

6 Blasch E, Breton R, Valin P, et al. User information fusion decision making analysis with the C-OODA model// 14th International Conference on Information Fusion. Chicago: IEEE, 2011: 1-8.

7 Shahbazian E, Blodgett D, Labbé P. The extended OODA model for data fusion systems// Proceedings of 2001 International Conference on Information Fusion. Montreal: International Conference on Information Fusion, 2001: 106-112.

8 Liggins M E, Hall D L, Llinas J. Handbook of Multisensor Data Fusion: Theory and Practice (Second Edition). New York: CRC Press, 2008.

9 Kuznetsova P, Ordonez V, Berg T L, et al. TreeTalk: Composition and compression of trees for image descriptions. Transactions of the Association for Computational Linguistics, 2014, 2: 351-362.

10 Rövid A, Remeli V. Towards raw sensor fusion in 3D object

detection// 2019 IEEE 17th World Symposium on Applied Machine Intelligence and Informatics (SAMI). Herlany: IEEE, 2019: 293-298.

11 Lathuilière S, Massé B, Mesejo P, et al. Deep reinforcement learning for audio-visual gaze control// 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Madrid: IEEE, 2019: 1555-1562.

12 Liang X C, Hu P Y, Zhang L G, et al. MCFNet: Multi-layer concatenation fusion network for medical images fusion. IEEE Sensors Journal, 2019, 19(16): 7107-7119.

13 Laghmara H, Laurain T, Cudel C, et al. Heterogeneous sensor data fusion for multiple object association using belief functions. Information Fusion, 2020, 57: 44-58.

14 李爱华, 续维佳, 石勇 . 基于数据融合的商务智能与分析架构研究 . 计算机科学, 2022, 49(12): 185-194.

Li A H, Xu W J, Shi Y. Framework of business intelligence and analysis based on data fusion. Computer Science, 2022, 49(12): 185-194. (in Chinese)

15 Gu J F, Zhu Z C. Knowing Wuli, sensing Shili, caring for Renli: Methodology of the WSR approach. Systemic Practice and Action Research, 2000, 13(1): 11-20.

16 朱志昌 . 物理事理人理方法论国际交流的启示// 中国系统工程学会第十一届学术年会论文集 . 宜昌: 中国系统工程学会, 2000: 149-164.

Zhu S C. The WSR approach in the international systems community// Proceeding of 11th Annual Conference of Systems Engineering Society of China. Yichang: Systems Engineering Society of China, 2000:149-164.(in Chinese)

17 顾基发, 唐锡晋 . 从古代系统思想到现代东方系统方法论 . 系统工程理论与实践, 2000, (1): 90-93.

Gu J F, Tang X J. From ancient system thoughts to modern oriental systems methodology. Systems Engineering - Theory & Practice, 2000, (1): 90-93. (in Chinese)

18 Li A H, Xu W J, Shi Y. A new data fusion framework of business intelligence and analytics in economy, finance and management// 2020 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT). New York: IEEE, 2020: 940-945.

19 彭怡, 寇纲 . 基于领域知识的数据挖掘理论框架研究// 第三届(2008)中国管理学年会——信息管理分会场论文集 . 长沙: 中国管理现代化研究会, 2008: 43-51.

Peng Y, Kou G. Research on data mining theory framework based on domain knowledge// The 3rd (2018) Chinese Academy Of Management Annual Conference- Collected Papers of Information Management Branch. Changsha: Chinese Academy of Management, 2008: 43-51. (in Chinese)

Research on multi-source heterogeneous big data fusion based on WSR

LI Aihua[1] XU Weijia[1] SHI Yong[2,3,4*]

(1 School of Management Science and Engineering, Central University of Finance and Economics, Beijing 100081, China;

2 School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190, China;

3 CAS Research Center on Fictitious Economy and Data Science, Beijing 100190, China;

4 Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences, Beijing 100190, China)

Abstract In the era of multi-source heterogeneous big data, big data presents new features such as cross, diversity and variability. The applications of big data in a wider range of fields have new requirements for data fusion. Under this background, the connotation of data fusion is enriched and expanded. The generalized data fusion includes the fusion of data resources, the fusion of model

*Corresponding author

methods, and the fusion of decision-makers' knowledge and experience. This study analyzes the characteristics of multi-source heterogeneous data fusion at three different fusion levels: data level, information level and decision level, and discusses challenges for data fusion in storage, application and analysis technology, data management as well as value determination. What's more, corresponding suggestions are putted forward, which benefit for enterprises, government and other entities to effectively manage data resources and provide reference for more in-depth data fusion analysis.

Keywords multi-source heterogeneous, wuli-shili-renli (WSR), data fusion, big data

李爱华 中央财经大学管理科学与工程学院教授。主要研究领域:大数据管理与应用、优化与管理决策、金融科技与风险管理等。E-mail: aihuali@cufe.edu.cn

LI Aihua Ph.D. in management science and engineering, Professor of School of Management Science and Engineering, Central University of Finance and Economics (CUFE). Her research focuses on big data management & application, optimization & management decision, fintech and risk management, etc. E-mail: aihuali@cufe.edu.cn

石 勇 发展中国家科学院院士。中国科学院大学教授。中国科学院大数据挖掘与知识管理重点实验室主任,中国科学院虚拟经济与数据科学研究中心主任。主要研究方向为数据挖掘与知识管理。 E-mail: yshi@ucas.ac.cn

SHI Yong Professor and Doctoral Supervisor of University of Chinese Academy of Sciences. Director of the Key Laboratory of Big Data Mining and Knowledge Management, Chinese Academy of Sciences (CAS), Director of the CAS Research Center on Fictitious Economy and Data Science, and Fellow of the World Academy of Sciences for the advancement of science in developing countries (TWAS). The main research directions are data mining and knowledge management. E-mail: yshi@ucas.ac.cn

■责任编辑: 张帆